Requirements for this paper:

Multi-choice cards:		Non-programmable calculator:	X	Open book examination?	NO
Graph paper:		Laptop:			

EXAMINATION/TEST:	**TEST 2 (MEMO)**	**QUALIFICATION:**	**HonsBSc**
MODULE CODE:	**ITRI 613**	**DURATION:**	**2 hours**
MODULE DESCRIPTION:	**Databases**	**MAX:**	**50 Marks**
EXAMINER(S):	**Dr H. Mogale**	**DATE:**	**21-05-21**
		TIME:	**18:00**

TOTAL: 50

INSTRUCTIONS TO CANDIDATE

Answer ALL questions

Make sure you have clear understanding of all the instructions and questions

Closed Book Examination

CANDIDATES ARE NOT ALLOWED TO READ QUESTIONS UNTIL THEY ARE TOLD TO DO SO BY THE CHIEF INVIGILATOR

TOTAL: 50 Marks

1. _____ is the mathematical basis for performing queries against a relational database.

- A. Linear Algebra
- B. Relational Algebra
- C. Relational Logic
- D. Relational Queries

[2 marks]

2. _____ stages pages from external storage to main memory buffer pool.

- A. Memory manager
- B. Heap
- C. Buffer manager
- D. Index manager

[2 marks]

3. The _____ operator of relational algebra assumes the elimination the duplicates.

- A. Natural Join
- B. Selection
- C. Union
- D. None of the above

[2 marks]

4. In SQL _____ is an optional keyword indicating that the answer should not contain duplicates.

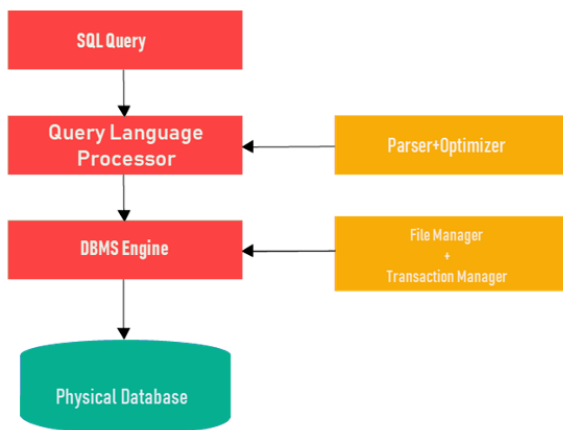- A. SELECT
- B. DISTINCT
- C. VIEW
- D. None of the above

[2 marks]

5. A _____ is a list of attributes of relations in relation-list.

- A. relation list
- B. query
- C. target-list
- D. qualification

[2 marks]

6. **Using the given diagram explain how an SQL query is processed and how the relevant data is retrieved from the database?**

Answer:

SQL processing is the parsing, optimization, row source generation, and execution of a SQL statement. Depending on the statement, the database may omit some of these stages.

The first stage of SQL processing is parsing. The parsing stage involves separating the pieces of a SQL statement into a data structure that other routines can process. The database parses a statement when instructed by the application, which means that only the application, and not the database itself, can reduce the number of parses. During execution stage, the SQL engine executes each row source in the tree produced by the row source generator. The database during execution stage reads the data from disk into memory if the data is not in memory. The database also takes out any locks and latches necessary to ensure data integrity and logs any changes made during the SQL execution. The final stage of processing a SQL statement is closing the cursor.
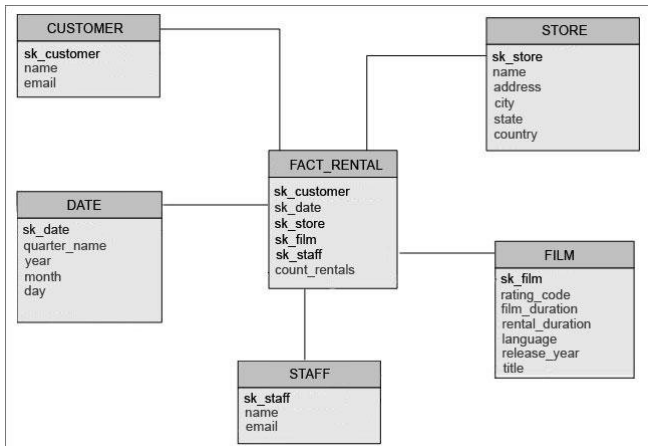
[4 marks]

7. Given the following schema write the following SQL queries?

1. Select all attributes from FACT_RENTAL?

2. Drop the table STORE?

3. Create a table LUXURY_VEHICLES which inherits all attributes of CUSTOMER and FACT_RENTAL.

ANSWER:

```
1. SELECT * FROM FACT_RENTAL;

2. DROP TABLE STORE;

3. CREATE TABLE LUXURY_VEHICLES (
      sk_cutomer INTEGER,
      name NVARCHAR(255),
      email NVARCHAR(255),
      sk_date DATE,
      sk_store NVARCHAR(255),
      sk_flim NVARCHAR(255),
      sk_staff NVARCHAR(255),
      count_rentals INTEGER,
   PRIMARY KEY (sk_customer)
);
```

[8 marks]

8. **In your own words and using examples kindly provide the primary differences in terms of uses for Heaps, Sorted files and Indexes?**

    Answer:

- **Heap files:**

    A heap file stores records of a file in a collection of disk pages in random order. Heap files are useful in scan operations where you step through all of the records in a file one at a time.

    A good example would be when you have an unordered relation called **Student** with attributes such as **first_name, last_name and student_num**. There is no ordering of the records in this example, which makes it a heap file.

- **Sorted file:**

    A sorted file is a file in which the records within the file are sorted according to some particular order.

    A good example when you have a relation called **Student** with attributes such as **first_name**, **last_name** and **student_num** and you order the records within the student relation alphabetically on the attribute "**last_name**".

- **Index:**

    An index is a data structure that uses search key values to organize records in a hierarchical tree-like structure or in a bucket (group containing equivalent values).

    A good example would be if you have the same student relation and implement a hash-based index on the attribute "last_name" and it returns all records related to a specific last name.

    [6 marks]

9. **Using an example provide the difference between an Equality query and a Range query?**

    ANSWER:

    Equality Query - This type of query is specifically designed to be able to find records that meet a specific criteria. Also known as an equality selection.
   Example - A query that needs to find all students who are 20yrs old and have average grade of 75.
    age = 20 AND grade = 75

    Range Query - This type of query is specifically designed to find records that satisfies a specific range.
    Example - A query that needs to find all students who have obtained a distinction for their grade.
    grade = > 75

    [4 marks]

10. **Given the following schema what will be the equivalent relational algebra to retrieve all employees with salaries above 15000...?**
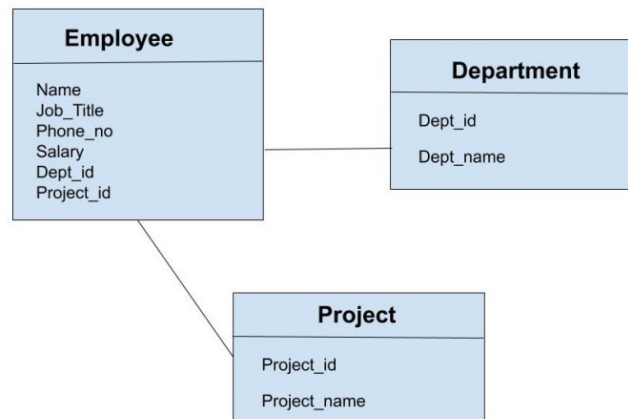
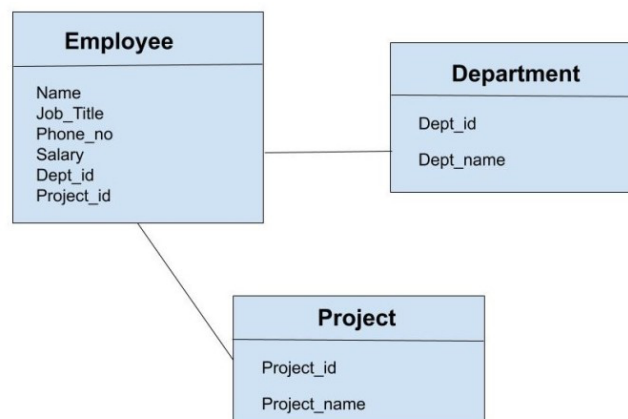Schema

A. σ Salary = 15000 (Employee)
B. σ Salary = 15000 (schema)
C. σ Salary > 15000 (Employee)
D. π Salary (σ Salary > 15000 (Employee))

[4 marks]

11. **Given the following schema what will be the equivalent relational algebra to retrieve only employees from finance department...?**

Schema

A. σ Employee = 'Finance' (Department)
B. π Employee = 'Finance' (Department)
C. σ Dept_id = 'Finance' (Employee)
D. π Dept_id = 'Finance' (Employee)

[4 marks]

12. **Assuming there are no errors returned kindly explain what the following SQL query returns after a successful execution? (Explain line by line)**

```
SELECT  S.rating,  MIN (S.age)
                            AS minage
FROM  Sailors S
WHERE  S.age >= 18
GROUP BY  S.rating
HAVING  COUNT (*) > 1
```

[4 marks]

Answer:

When the code executes it will find the age of the youngest sailor with age 18, for each rating which has at least 2 such sailors.

Line 1: `SELECT S.rating, MIN(S.age) AS minage:`

Specifies the target-list.

Line 2: `From Sailors S:`

Specifies the relation-list. The "Sailor" relation is specified and denoted with the range variable "S".

Line 3: `WHERE S.age >= 18:`

Specifies the qualification. The qualification specifies that only tuples from the "Sailor" relation in which the "age" attribute is greater or equal to 18 should be considered in the result of the query.

Line 4: `GROUP BY S.rating:`

Specifies the grouping-list. Groups the results of the query by rating. Since we are using an aggregate operation in the SELECT clause as well as selecting the "rating", we need to implement grouping for the SQL query to be valid.

Line 5: `HAVING COUNT(*) > 1:`

Specifies the grouping-qualification. The grouping qualification specifies that only sailors with rating that is at least two and above should be returned.

13. **With the help of an example kindly discuss the primary differences between Clustered and Unclustered index classifications?**

    Answer:

    Cluster index is a type of index which sorts the data rows in the table on their key values. In the Database, there is only one clustered index per table. A clustered index defines the order in which data is stored in the table which can be sorted in only one way. So, there can be an only a single clustered index for every table.

    A Non-clustered index stores the data at one location and indices at another location. The index contains pointers to the location of that data. A single table can have many non-clustered indexes as an index in the non-clustered index is stored in different places.

    For example, a book can have more than one index, one at the beginning which displays the contents of a book unit wise while the second index shows the index of terms in alphabetical order.

    Cluster index is a type of index that sorts the data rows in the table on their key values whereas the Non-clustered index stores the data at one location and indices at another location.

    Clustered index stores data pages in the leaf nodes of the index while Non-clustered index method never stores data pages in the leaf nodes of the index.

    [4 marks]

14. **Name any two of RDBMS packages of your choosing and discuss their unique features and characteristics and also explains which real life companies are using them today?**

    Answer:

    MySQL - is the popular database management system used for managing the relational database. It is open-source database software, which is supported by Oracle Company. It is fast, scalable, and easy to use database management system in comparison with Microsoft SQL Server and Oracle Database. It is commonly used with PHP scripts for creating powerful and dynamic server-side or web-based enterprise applications.

    Many companies today still use MySQL including coca-cola, MTN, RMB etc.

    IBM Db2 - Database on IBM Cloud Pak for Data combines is a proven, AI-infused, enterprise-ready data management system with an integrated data and AI platform built on the security-rich, scalable Red Hat® OpenShift® foundation. Derive insights with machine learning embedded into query processing. Cut costs with the multimodal capability that eliminates the need for data replication and migration.

    Many companies today use DB2 including IBM, Apple, Morgan Stainley, John Deere etc.

    [2 marks]

    [Total – 50 Marks]